

PF-0501-1 DIV

**PROSTATE ASSOCIATED ETS PROTEIN**

This application is a divisional application of USN 09/055,113, filed April 4, 1998, the contents of which are hereby incorporated by reference.

5

**FIELD OF THE INVENTION**

This invention relates to nucleic acid and amino acid sequences of a prostate associated Ets protein and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, immune, reproductive, and developmental disorders.

10

**BACKGROUND OF THE INVENTION**

Ets proteins comprise a large, conserved family of transcription factors that regulate the expression of genes involved in growth control, cell differentiation, and development. Over 30 Ets proteins have been identified in higher eukaryotes ranging from the fruit fly Drosophila melanogaster to humans. Ets proteins are defined by a unique DNA-binding domain, called the Ets domain, which is structurally and functionally distinct from DNA-binding domains in other transcription factors. Ets proteins are encoded by proto-oncogenes which, when misregulated, contribute to the development and progression of neoplastic diseases. (See reviews by MacLeod, K. et al. (1992) Trends Biochem. Sci. 17:251-256; Seth, A. et al. (1992) Cell Growth Differ. 3:327-334; and Wasyluk, B. et al. (1993) Eur. J. Biochem. 211:7-18.)

15

20

25

30

The founding member of the *ets* gene family is the *v-ets* oncogene from the acutely transforming avian retrovirus, E26, which causes erythroid and myeloid leukemia in chickens. *v-ets* (E26 transformation-specific) is derived from the *c-ets-1* proto-oncogene, which is normally present in the chicken genome. Integration of this proto-oncogene into the viral genome resulted in the formation of an oncogenic fusion gene containing *v-ets*. Since the discovery of *v-ets* and its progenitor, *c-ets-1*, many *ets*-related proto-oncogenes have been characterized in diverse species, with up to several divergent *ets* genes present in a single species. These divergent genes appear to have been generated by multiple gene duplication events involving a single ancestral *ets* gene.

All Ets proteins contain a characteristic Ets domain, which is necessary and sufficient for DNA binding. This domain encompasses about 85 amino acids located in the C-terminal region of the protein. The C-terminal portion of this domain is enriched in basic amino acid residues that may directly interact with DNA. The Ets domain also contains three diagnostic  
5 tryptophan residues, spaced about 17 to 19 amino acids apart, and a nuclear localization signal. Structural analyses predict that regions of the Ets domain surrounding the first and third tryptophans are  $\alpha$ -helical. The Ets domain can bind to DNA as a monomer, unlike other DNA-binding domains which require oligomerization for DNA-binding activity. In addition, some Ets proteins also contain an N-terminal helix-loop-helix domain, which may be  
10 involved in protein-protein interactions. This domain, together with the remainder of the protein, is conserved weakly among Ets proteins.

The Ets domain binds to a 10-nucleotide motif with an invariant purine trinucleotide core. Variations in the nucleotides flanking the core may contribute to the specificity of Ets binding. The Ets-binding motif has been identified in both the 5' and 3' regulatory regions of  
15 a variety of genes, many of which are involved in T-cell differentiation and proliferation. Such genes include, for example, those encoding T-cell receptors, cytokines, hematopoietic growth factors, and other proteins involved in T-cell signaling. In addition, some proto-oncogenes such as *fos* and the *ets* genes themselves contain Ets-binding motifs in their promoters. Interestingly, Ets-binding motifs are also present in the regulatory regions of  
20 some viral genomes, including those of human immunodeficiency virus-1 and Epstein-Barr virus.

*ets* gene expression is responsive to mitogens that trigger Ras-, calcium-, and protein kinase C-mediated signal transduction pathways. In addition, Ets protein activity is regulated by post-translational modifications and protein-protein interactions. For example, two  
25 human Ets proteins, ETS1 and ETS2, are inactivated by phosphorylation in differentiated T-cells. Moreover, the interaction of ETS1 and ETS2 with other transcription factors, such as Fos and Jun, determines their DNA-binding specificity and level of activity. (Basuyaux, J. P. et al. (1997) J. Biol. Chem. 272:26188-26195.)

Ets proteins are important for various developmental processes. In Drosophila, seven  
30 different Ets proteins are expressed at various times during development. One of these

proteins, E74A, is specifically involved in the metamorphosis from larva to adult. Another protein, Yan/Pok, plays a role in the formation of photoreceptor cells in the developing eye. A third protein, encoded by D-*ets*-4, is expressed in pole cells, which ultimately give rise to the adult germline, during a restricted time period of early embryonic development. (Chen, T. et al. (1992) Dev. Biol. 151:176-191.) In mice, the *ets* gene, *TEL*, is essential for normal development. Homozygous deletion of *TEL* is lethal during early embryogenesis. Blood vessel formation is defective in the extra-embryonic yolk sac, and neural and mesenchymal cells undergo apoptosis within the embryo. (Wang, L. C. et al. (1997) EMBO 16:4374-4383; Golub, T. et al. (1994) Cell 77:307-316.) In addition, human *ETS1* and *ETS 2* appear to play a distinct role in T-cell maturation in the fetal thymus.

The conversion of *ets* proto-oncogenes to oncogenes with transforming activity can be achieved by three basic mechanisms: overexpression, proviral integration, or chromosomal translocation. First, overexpression of either human *ETS1* or *ETS2* transforms NIH3T3 mouse fibroblasts. The transformed cells induce tumors when injected into immunodeficient mice. In addition, individuals with Down's syndrome, caused by an extra copy of all or part of chromosome 21, are predisposed to develop leukemia. The *ets* genes *ERG* and *ETS2* are located in the minimal region of chromosome 21 associated with Down's syndrome, suggesting that increased dosage of *ERG* and *ETS2* contributes to leukemogenesis. (Papas, T. S. et al. (1990) Am. J. Med. Genet. Suppl. 7:251-261.) Second, virus-mediated erythroleukemia is induced in mice by integration of Friend murine leukemia provirus or spleen focus-forming provirus into the host genome adjacent to *ets* coding regions. The viral regulatory sequences act as enhancers of host *ets* expression. Third, chromosomal translocations involving *ets* loci have been implicated in several leukemias and lymphomas. For example, a chromosomal translocation which fuses the region of *TEL* encoding the Ets domain to the *MNI* gene results in the expression of a chimeric transcription factor implicated in myeloid leukemias. (Buijs, A. et al. (1995) Oncogene 10:1511-1519.)

*ets* gene activity may also contribute to the metastasis of existing tumors by promoting tumor vascularization and by stimulating the expression of enzymes that break down extracellular matrix proteins. Both of these activities allow tumor cells to migrate from the site of the primary tumor.

The discovery of a new prostate associated Ets protein and the polynucleotides encoding it satisfies a need in the art by providing new compositions which are useful in the diagnosis, treatment, and prevention of cell proliferative, immune, reproductive, and developmental disorders.

5

## SUMMARY OF THE INVENTION

The invention is based on the discovery of a new human prostate associated Ets protein (PRAEP), the polynucleotides encoding PRAEP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, immune, reproductive, and developmental disorders. The invention features a substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention further provides a substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention further provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, as well as an isolated and purified polynucleotide which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides an isolated and purified polynucleotide comprising the polynucleotide sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2, and an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide comprising the polynucleotide sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2. The invention also provides an isolated and purified polynucleotide having a sequence complementary to the polynucleotide comprising the polynucleotide sequence of

SEQ ID NO:2 or a fragment of SEQ ID NO:2.

The invention further provides an expression vector comprising at least a fragment of the polynucleotide encoding the polypeptide comprising the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, the method comprising the steps of: (a) culturing the host cell comprising an expression vector containing at least a fragment of a polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO: 1 under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide comprising the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, as well as a purified agonist and a purified antagonist of the polypeptide.

The invention also provides a method for treating or preventing a cell proliferative disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing an immune disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing a reproductive disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing a developmental disorder, the method comprising administering to a subject in need of such treatment an

effective amount of an antagonist of the polypeptide having the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for detecting a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO: 1 or a fragment of SEQ ID NO: 1 in a biological sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 to at least one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 in the biological sample. In one aspect, the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.

## BRIEF DESCRIPTION OF THE FIGURES

Figures 1A, 1B, 1C, 1D, and 1E show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of PRAEP. The alignment was produced using MACDNASIS PRO software (Hitachi Software Engineering Co. Ltd., San Bruno, CA).

Figure 2 shows the amino acid sequence alignments among the Ets domains of PRAEP (1813005; SEQ ID NO:1), Drosophila D-ets-4 (GI 157196; SEQ ID NO:3), and human TEL (GI 511283; SEQ ID NO:4), produced using the multisequence alignment program of LASERGENE software (DNASTAR Inc, Madison WI).

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to “a host cell” includes a plurality of such host cells, and a reference to “an antibody” is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

#### DEFINITIONS

“PRAEP,” as used herein, refers to the amino acid sequences of substantially purified PRAEP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term “agonist,” as used herein, refers to a molecule which, when bound to PRAEP, increases or prolongs the duration of the effect of PRAEP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of PRAEP.

An “allelic variant,” as this term is used herein, is an alternative form of the gene encoding PRAEP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times

in a given sequence.

"Altered" nucleic acid sequences encoding PRAEP, as described herein, include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as PRAEP or a polypeptide with at least one functional characteristic of PRAEP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PRAEP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PRAEP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PRAEP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PRAEP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of PRAEP which are preferably about 5 to about 15 amino acids in length, most preferably 14 amino acids, and which retain some biological activity or immunological activity of PRAEP. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art. (See, e.g., Dieffenbach, C.W. and G.S. Dveksler



(1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5.)

The term "antagonist," as it is used herein, refers to a molecule which, when bound to PRAEP, decreases the amount or the duration of the effect of the biological or immunological activity of PRAEP. Antagonists may include proteins, nucleic acids, carbohydrates,

5 antibodies, or any other molecules which decrease the effect of PRAEP.

As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind PRAEP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing

10 antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

15 The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may

20 compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or

25 transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural,

30 regulatory, or biochemical functions of a naturally occurring molecule. Likewise,

"immunologically active" refers to the capability of the natural, recombinant, or synthetic PRAEP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence," as these terms are used herein, refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation, an aqueous solution, or a sterile composition. Compositions comprising polynucleotide sequences encoding PRAEP or fragments of PRAEP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts, e.g., NaCl, detergents, e.g., sodium dodecyl sulfate (SDS), and other components, e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.

"Consensus sequence," as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR amplification kit (Applied Biosystems, Foster City, CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence.

As used herein, the term "correlates with expression of a polynucleotide" indicates

that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding PRAEP, by Northern analysis is indicative of the presence of nucleic acids encoding PRAEP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding PRAEP.

5           A "deletion," as the term is used herein, refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

          The term "derivative," as used herein, refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a  
10   polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

15           The term "similarity," as used herein, refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the  
20   target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific  
25   binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize  
30   to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Inc., Madison WI). The MEGALIGN program can create alignments between two or more sequences according to different methods, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) *Gene* 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) *Methods Enzymol.* 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 Kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

The term "humanized antibody," as used herein, refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization," as the term is used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t

analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

5           The words “insertion” or “addition,” as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

          “Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be  
10       characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

          The term “microarray,” as used herein, refers to an arrangement of distinct polynucleotides arrayed on a substrate, e.g., paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

15           The terms “element” or “array element” as used herein in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

          The term “modulate,” as it appears herein, refers to a change in the activity of PRAEP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PRAEP.

20           The phrases “nucleic acid” or “nucleic acid sequence,” as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, “fragments” refers to those  
25       nucleic acid sequences which, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

          The terms “operably associated” or “operably linked,” as used herein, refer to functionally related nucleic acid sequences. A promoter is operably associated or operably  
30       linked with a coding sequence if the promoter controls the translation of the encoded

polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

5           The term “oligonucleotide,” as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. As used herein, the term “oligonucleotide” is substantially equivalent to the terms “amplimer,” “primer,” “oligomer,” and “probe,” as these terms are  
10           commonly defined in the art.

          “Peptide nucleic acid” (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single  
15           stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell. (See, e.g., Nielsen, P.E. et al. (1993) *Anticancer Drug Des.* 8:53-63.)

          The term “sample,” as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding PRAEP, or fragments thereof, or PRAEP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or  
20           membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

          As used herein, the terms “specific binding” or “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the  
25           antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

          As used herein, the term “stringent conditions” refers to conditions which permit  
30           hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions

can be defined by salt concentration, the concentration of organic solvent (e.g., formamide), temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

5 For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least  
10 about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to  
15 those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred  
20 embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above,  
25 wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least  
30 about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash

steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these  
5 conditions will be readily apparent to those skilled in the art.

The term “substantially purified,” as used herein, refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

10 A “substitution,” as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

“Transformation,” as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known  
15 method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term “transformed” cells includes stably transformed cells in which the inserted DNA is capable of replication either as an  
20 autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A “variant” of PRAEP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have “conservative” changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of  
25 leucine with isoleucine). More rarely, a variant may have “nonconservative” changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for  
30 example, LASERGENE software.



## THE INVENTION

The invention is based on the discovery of a new human prostate associated Ets protein (PRAEP), the polynucleotides encoding PRAEP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, immune, reproductive, and developmental disorders.

Nucleic acids encoding the PRAEP of the present invention were first identified in Incyte Clone 1813005 from the prostate tumor cDNA library (PROSTUT12) using a computer search, e.g., BLAST, for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 1813005 and 1810124 (PROSTUT12), 1995317 (BRSTTUT03), and 581952 (PROSNOT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1, as shown in Figures 1A, 1B, 1C, 1D, and 1E. PRAEP is 335 amino acids in length and has one potential N-glycosylation site at N142; ten potential casein kinase II phosphorylation sites at T26, S61, S81, S82, T113, S229, T230, S234, T236, and S238; three potential protein kinase C phosphorylation sites at S81, S308, and S324; and two potential calmodulin-dependent protein kinase II phosphorylation sites at S44 and S195. Protein sequence analysis using various search algorithms indicates that PRAEP contains an Ets domain and is therefore a member of the Ets protein family. Motifs analysis indicates that PRAEP contains two Ets domain signatures corresponding to two different conserved regions of the Ets domain. The first signature extends from L251 to L259, and the second extends from R297 to Y312. BLOCKS analysis also indicates that the region of PRAEP from I278 to V328 shows significant homology to a conserved region of the Ets domain. PRINTS analysis further indicates that amino acid sequences in PRAEP from I249 to P262 and from E275 to V332 are homologous to four distinct Ets domain signatures. Likewise, Hidden Markov Model analysis indicates that the region of PRAEP from I249 to F331 is homologous to an Ets domain consensus sequence. Furthermore, the three conserved tryptophan residues characteristic of Ets domains are conserved in PRAEP at W252, W271, and W291. Secondary structure predictions indicate that residues surrounding these tryptophans likely form  $\alpha$ -helices. In addition, the isoelectric point of the region of PRAEP comprising the predicted Ets domain from I249 to V332 is 10.6, indicating that this region is

enriched in basic amino acids. As shown in Figure 2, the predicted Ets domain of PRAEP has chemical and structural similarity with the Ets domains of Drosophila D-ets-4 (GI 157196; SEQ ID NO:3) and of human TEL (GI 511283; SEQ ID NO:4). In particular, the Ets domain of PRAEP shares 75% identity with that of D-ets-4 and 44% identity with that of TEL. Note that in Figure 2, only the Ets domains of the three proteins are shown, and the amino acid residues are numbered based on their positions in the complete protein sequences as reported in the Sequence Listing. Two of the potential protein kinase C phosphorylation sites in the PRAEP Ets domain are conserved in D-ets-4. A region of unique sequence in PRAEP from about amino acid 7 to about amino acid 18 is encoded by a fragment of SEQ ID NO:2 from about nucleotide 448 to about nucleotide 483. Northern analysis shows the expression of this sequence in various libraries, at least 87% of which are involved in cancer and cell proliferation. In particular, 82% of the libraries expressing PRAEP are derived from reproductive tissue: 47% are derived specifically from prostate tissue, and 33% are derived specifically from breast tissue.

The invention also encompasses PRAEP variants. A preferred PRAEP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the PRAEP amino acid sequence, and which contains at least one functional or structural characteristic of PRAEP.

The invention also encompasses polynucleotides which encode PRAEP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:2, which encodes an PRAEP.

The invention also encompasses a variant of a polynucleotide sequence encoding PRAEP. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PRAEP. A particular aspect of the invention encompasses a variant of SEQ ID NO:2 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:2. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PRAEP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PRAEP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PRAEP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PRAEP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring PRAEP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PRAEP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PRAEP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PRAEP and PRAEP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PRAEP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:2, or a fragment of SEQ ID NO:2, under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.)

Methods for DNA sequencing are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE DNA polymerase (US Biochemical Corp., Cleveland, OH), Taq polymerase (Applied Biosystems), 5 thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (GIBCO BRL, Gaithersburg, MD). Preferably, the process is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno, NV), Peltier thermal cycler (PTC200; MJ Research, Watertown, MA) and the ABI CATALYST and 373 and 377 DNA 10 sequencers (Applied Biosystems).

The nucleic acid sequences encoding PRAEP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify 15 unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A 20 third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to 25 retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries to walk genomic DNA (Clontech, Palo Alto, CA). This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, 30 such as OLIGO 4.06 primer analysis software (National Biosciences Inc., Plymouth, MN) or

another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often  
5 include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In  
10 particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of  
15 samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PRAEP may be cloned in recombinant DNA molecules that direct  
20 expression of PRAEP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PRAEP.

The nucleotide sequences of the present invention can be engineered using methods  
25 generally known in the art in order to alter PRAEP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may  
30 be used to introduce mutations that create new restriction sites, alter glycosylation patterns,

change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding PRAEP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res.

5 Symp. Ser. 225-232.) Alternatively, PRAEP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.)

Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PRAEP, or any part thereof, may be  
10 altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol.

15 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman and Co., New York, NY.)

In order to express a biologically active PRAEP, the nucleotide sequences encoding PRAEP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of  
20 the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PRAEP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PRAEP. Such signals include the ATG initiation  
25 codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PRAEP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon  
30 should be provided by the vector. Exogenous translational elements and initiation codons

may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct  
 5 expression vectors containing sequences encoding PRAEP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, ch. 4, 8, and 16-17; and Ausubel, F.M. et al. (1995, and periodic supplements) Current  
 10 Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PRAEP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with  
 15 viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected  
 20 depending upon the use intended for polynucleotide sequences encoding PRAEP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PRAEP can be achieved using a multifunctional E. coli vector such as BLUESCRIPT phagemid (Stratagene) or PSPORT1 plasmid (GIBCO BRL). Ligation of sequences encoding PRAEP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric  
 25 screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of PRAEP are needed, e.g. for the production of  
 30 antibodies, vectors which direct high level expression of PRAEP may be used. For example,

vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PRAEP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, supra; and Grant et al. (1987) *Methods Enzymol.* 153:516-54; Scorer, C. A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of PRAEP. Transcription of sequences encoding PRAEP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) *EMBO J.* 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PRAEP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PRAEP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes,



polycationic amino polymers, or vesicles) for therapeutic purposes.

For long term production of recombinant proteins in mammalian systems, stable expression of PRAEP in cell lines is preferred. For example, sequences encoding PRAEP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences.

Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*<sup>-</sup> or *apr*<sup>-</sup> cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; and Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP) (Clontech, Palo Alto, CA),  $\beta$  glucuronidase and its substrate  $\beta$ -D-glucuronoside, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed.

For example, if the sequence encoding PRAEP is inserted within a marker gene sequence, transformed cells containing sequences encoding PRAEP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PRAEP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PRAEP and that express PRAEP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PRAEP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PRAEP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Section IV; Coligan, J. E. et al. (1997 and periodic supplements) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York, NY; and Maddox, D.E. et al. (1983) J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PRAEP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PRAEP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available

kits, such as those provided by Pharmacia (Kalamazoo, MI), Promega (Madison, WI), and U.S. Biochemical Corp. (Cleveland, OH). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PRAEP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PRAEP may be designed to contain signal sequences which direct secretion of PRAEP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Manassas, VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PRAEP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PRAEP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PRAEP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and

hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PRAEP encoding sequence and the heterologous protein sequence, so that PRAEP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel, F. M. et al. (1995 and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch 10. A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PRAEP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract systems (Promega, Madison, WI). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably <sup>35</sup>S-methionine.

Fragments of PRAEP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using an Applied Biosystems 431A peptide synthesizer (Applied Biosystems). Various fragments of PRAEP may be synthesized separately and then combined to produce the full length molecule.

## **THERAPEUTICS**

Chemical and structural similarity exists among PRAEP, D-ets-4 from Drosophila (GI 157196), and TEL from human (GI 511283). In addition, PRAEP is expressed in reproductive tissue, particularly in the prostate. Therefore, PRAEP appears to play a role in cell proliferative, immune, reproductive, and developmental disorders.

Therefore, in one embodiment, an antagonist of PRAEP may be administered to a subject to treat or prevent a cell proliferative disorder. Such a disorder may include, but is

not limited to, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds PRAEP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express PRAEP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PRAEP may be administered to a subject to treat or prevent a cell proliferative disorder including, but not limited to, those described above.

In another embodiment, an antagonist of PRAEP may be administered to a subject to treat or prevent an immune disorder. Such a disorder may include, but is not limited to, acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome,

complications of cancer, hemodialysis, and extracorporeal circulation, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease. In one aspect, an antibody which specifically binds PRAEP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which  
5 express PRAEP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PRAEP may be administered to a subject to treat or prevent an immune disorder including, but not limited to, those described above.

In a further embodiment, an antagonist of PRAEP may be administered to a subject to  
10 treat or prevent a reproductive disorder. Such a disorder may include, but is not limited to, abnormal prolactin production, infertility, tubal disease, ovulatory defects, endometriosis, perturbations of the estrous and menstrual cycles, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, teratogenesis, breast cancer, fibrocystic breast disease, galactorrhea, abnormal  
15 spermatogenesis, abnormal sperm physiology, testicular cancer, prostate cancer, benign prostatic hyperplasia, prostatitis, and gynecomastia. In one aspect, an antibody which specifically binds PRAEP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express PRAEP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PRAEP may be administered to a subject to treat or prevent a  
20 reproductive disorder including, but not limited to, those described above.

In still another embodiment, an antagonist of PRAEP may be administered to a subject to treat or prevent a developmental disorder. The term "developmental disorder"  
25 refers to any disorder associated with growth and differentiation, embryogenesis, and morphogenesis involving any tissue, organ, or system of a subject (such as the brain, adrenal gland, kidney, skeletal or reproductive system). Such a disorder may include, but is not limited to, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome  
30 (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary

keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spinal bifida, congenital glaucoma, cataract, and sensorineural hearing loss. In one aspect, an antibody which specifically binds PRAEP may be used  
5 directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express PRAEP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PRAEP may be administered to a subject to treat or prevent a developmental disorder including, but not limited to, those described above.

10 In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act  
15 synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PRAEP may be produced using methods which are generally known in the art. In particular, purified PRAEP may be used to produce antibodies or to screen  
20 libraries of pharmaceutical agents to identify those which specifically bind PRAEP. Antibodies to PRAEP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for  
25 therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PRAEP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include,  
30 but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH,

and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PRAEP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of PRAEP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PRAEP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PRAEP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; and Winter, G. et al. (1991) Nature 349:293-299.)



Antibody fragments which contain specific binding sites for PRAEP may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression  
 5 libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric  
 10 assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PRAEP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PRAEP epitopes is preferred, but a competitive binding assay may also be employed. (Maddox,  
 15 supra.)

In another embodiment of the invention, the polynucleotides encoding PRAEP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding PRAEP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be  
 20 transformed with sequences complementary to polynucleotides encoding PRAEP. Thus, complementary molecules or fragments may be used to modulate PRAEP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PRAEP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding PRAEP. (See, e.g., Sambrook, supra; and  
 30 Ausubel, supra.)

Genes encoding PRAEP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding PRAEP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may  
5 continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or  
10 regulatory regions of the gene encoding PRAEP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory  
15 molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

20 Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding  
25 PRAEP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing  
30 the cleavage site, may be evaluated for secondary structural features which may render the

oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PRAEP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable

carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of PRAEP, antibodies to PRAEP, and mimetics, agonists, antagonists, or inhibitors of PRAEP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any  
5 sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular,  
10 intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used  
15 pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral  
20 administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally,  
25 after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin  
30 and collagen. If desired, disintegrating or solubilizing agents may be added, such as the

cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of PRAEP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PRAEP or fragments thereof, antibodies of PRAEP, and agonists, antagonists or inhibitors of PRAEP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity.

The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind PRAEP may be used for the diagnosis of disorders characterized by expression of PRAEP, or in assays to monitor patients being treated with PRAEP or agonists, antagonists, or inhibitors of PRAEP.

Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PRAEP include methods which utilize the antibody and a label to detect PRAEP in human body fluids or in extracts of cells or tissues.

The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PRAEP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PRAEP expression. Normal or standard values for PRAEP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with

antibody to PRAEP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of PRAEP expressed in subject samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PRAEP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of PRAEP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PRAEP, and to monitor regulation of PRAEP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PRAEP or closely related molecules may be used to identify nucleic acid sequences which encode PRAEP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding PRAEP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the PRAEP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:2 or from genomic sequences including promoters, enhancers, and introns of the PRAEP gene.

Means for producing specific hybridization probes for DNAs encoding PRAEP include the cloning of polynucleotide sequences encoding PRAEP or PRAEP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides.



Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PRAEP may be used for the diagnosis of a disorder associated with expression of PRAEP. Examples of such a disorder include, but are not limited to, a cell proliferative disorder such as arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immune disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease; a reproductive disorder such

as abnormal prolactin production, infertility, tubal disease, ovulatory defects, endometriosis, perturbations of the estrous and menstrual cycles, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, teratogenesis, breast cancer, fibrocystic breast disease, galactorrhea, abnormal spermatogenesis, abnormal sperm physiology, testicular cancer, prostate cancer, benign prostatic hyperplasia, prostatitis, and gynecomastia; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spinal bifida, congenital glaucoma, cataract, and sensorineural hearing loss. The polynucleotide sequences encoding PRAEP may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered PRAEP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PRAEP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PRAEP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PRAEP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression

of PRAEP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PRAEP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PRAEP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding PRAEP, or a fragment of a polynucleotide complementary to the polynucleotide encoding PRAEP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of PRAEP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol.

Methods 159:235-244; and Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

5 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a  
10 disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application  
15 WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding PRAEP may be used to generate hybridization probes useful in mapping the naturally occurring  
20 genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet.  
25 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R.A. (ed.) Molecular Biology and Biotechnology, VCH Publishers New York, NY, pp. 965-968.) Examples of genetic map data can be found in various scientific  
30 journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between

the location of the gene encoding PRAEP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

5        In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to  
10        chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., AT to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988)  
15        Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

      In another embodiment of the invention, PRAEP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any  
20        of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PRAEP and the agent being tested may be measured.

      Another technique for drug screening provides for high throughput screening of  
25        compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with PRAEP, or fragments thereof, and washed. Bound PRAEP is then detected by methods well known in the art. Purified PRAEP can also be  
30        coated directly onto plates for use in the aforementioned drug screening techniques.

Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PRAEP specifically compete with a test compound for binding PRAEP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PRAEP.

In additional embodiments, the nucleotide sequences which encode PRAEP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

## EXAMPLES

### I. PROSTUT12 cDNA Library Construction

The PROSTUT12 cDNA library was constructed from prostate tumor tissue removed from a 65 year-old Caucasian male during a radical prostatectomy. The patient presented with elevated serum levels of prostate specific antigen and was diagnosed as having a malignant prostate neoplasm. Pathology indicated an adenocarcinoma (Gleason grade 2+2) involving the right anterior prostate peripherally. Multiple microscopic foci were identified in the left and right sides of the prostate but did not involve the capsule. Adenofibromatous hyperplasia and perineural invasion were present. Surrounding surgical margins and multiple pelvic lymph nodes were negative for tumor. Patient history included a transurethral prostatectomy.

The frozen tissue was homogenized and lysed in guanidinium isothiocyanate solution using a Polytron PT-3000 homogenizer (Brinkmann Instruments, Westbury, NY). The lysate was centrifuged over a CsCl cushion to isolate RNA. The RNA was extracted with acid phenol, precipitated with sodium acetate and ethanol, resuspended in RNase-free water, and treated with DNase. The RNA was re-extracted with acid phenol and reprecipitated with sodium acetate and ethanol. Poly(A+) RNA was isolated using the OLIGOTEX kit

PF-0501-1 DIV

(QIAGEN Inc, Chatsworth, CA).

Poly(A+) RNA was used for cDNA synthesis and construction of the PROSTUT12 cDNA library according to the recommended protocols in the SUPERSCRIP<sup>T</sup> plasmid system (Catalog #18248-013, Gibco/BRL). The cDNAs were fractionated on a  
5 SEPHAROSE CL4B column (Catalog #275105-01, Pharmacia), and those cDNAs exceeding 400 bp were ligated into pINCY 1 (Incyte). The recombinant plasmids were subsequently transformed into DH5 $\alpha$  competent cells (Catalog #18258-012, Gibco/BRL).

## II. Isolation and Sequencing of cDNA Clones

10 Plasmid DNA was released from the cells and purified using the R.E.A.L. PREP 96 plasmid kit (Catalog #26173, QIAGEN Inc). The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, Gibco/BRL) with carbenicillin at 25 mg/l and glycerol at 0.4%; 2) after the cultures were incubated for 19 hours, the cells were lysed with 0.3 ml of lysis buffer;  
15 and 3) following isopropanol precipitation, the plasmid DNA pellets were each resuspended in 0.1 ml of distilled water. The DNA samples were stored at 4°C.

The cDNAs were sequenced by the method of Sanger et al. (1975, J. Mol. Biol. 94:441f), using a MICROLAB 2200 liquid transfer system (Hamilton, Reno, NV) in combination with Peltier thermal cyclers (PTC200 from MJ Research, Watertown, MA) and  
20 Applied Biosystems 377 DNA sequencing systems, and the reading frame was determined.

## III. Similarity Searching of cDNA Clones and Their Deduced Proteins

The nucleotide sequences and/or amino acid sequences of the Sequence Listing were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II databases. These  
25 databases, which contain previously identified and annotated sequences, were searched for regions of similarity using BLAST (Basic Local Alignment Search Tool). (See, e.g., Altschul, S.F. (1993) J. Mol. Evol 36:290-300; and Altschul et al. (1990) J. Mol. Biol. 215:403-410.)

BLAST produced alignments of both nucleotide and amino acid sequences to  
30 determine sequence similarity. Because of the local nature of the alignments, BLAST was

especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms could have been used when dealing with primary sequence patterns and secondary structure gap penalties. (See, e.g., Smith, T. et al. (1992) *Protein Engineering* 5:35-51.) The sequences disclosed in this application have lengths of at least 49 nucleotides and have no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at  $10^{-25}$  for nucleotides and  $10^{-8}$  for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate (pri), rodent (rod), and other mammalian sequences (mam), and deduced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp), for similarity.

Additionally, sequences identified from cDNA libraries may be analyzed to identify those gene sequences encoding conserved protein motifs using an appropriate analysis program, e.g., BLOCKS. BLOCKS is a weighted matrix analysis algorithm based on short amino acid segments, or blocks, compiled from the PROSITE database. (Bairoch, A. et al. (1997) *Nucleic Acids Res.* 25:217-221.) The BLOCKS algorithm is useful for classifying genes with unknown functions. (Henikoff S. And Henikoff G.J., *Nucleic Acids Research* (1991) 19:6565-6572.) Blocks, which are 3-60 amino acids in length, correspond to the most highly conserved regions of proteins. The BLOCKS algorithm compares a query sequence with a weighted scoring matrix of blocks in the BLOCKS database. Blocks in the BLOCKS database are calibrated against protein sequences with known functions from the SWISS-PROT database to determine the stochastic distribution of matches. Similar databases such as PRINTS, a protein fingerprint database, are also searchable using the BLOCKS algorithm. (Attwood, T. K. et al. (1997) *J. Chem. Inf. Comput. Sci.* 37:417-424.) PRINTS is based on non-redundant sequences obtained from sources such as SWISS-PROT, GenBank, PIR, and NRL-3D.

The BLOCKS algorithm searches for matches between a query sequence and the



BLOCKS or PRINTS database and evaluates the statistical significance of any matches found. Matches from a BLOCKS or PRINTS search can be evaluated on two levels, local similarity and global similarity. The degree of local similarity is measured by scores, and the extent of global similarity is measured by score ranking and probability values. A score of 1000 or greater for a BLOCKS match of highest ranking indicates that the match falls within the 0.5 percentile level of false positives when the matched block is calibrated against SWISS-PROT. Likewise, a probability value of less than  $1.0 \times 10^{-3}$  indicates that the match would occur by chance no more than one time in every 1000 searches. Only those matches with a cutoff score of 1000 or greater and a cutoff probability value of  $1.0 \times 10^{-3}$  or less are considered in the functional analyses of the protein sequences in the Sequence Listing.

In another alternative, Hidden Markov models (HMMs) may be used to find protein domains, each defined by a dataset of proteins known to have a common biological function. (See, e.g., Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; and Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197.) HMMs were initially developed to examine speech recognition patterns, but are now being used in a biological context to analyze protein and nucleic acid sequences as well as to model protein structure. (See, e.g., Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; and Collin, M. et al. (1993) Protein Sci. 2:305-314.) HMMs have a formal probabilistic basis and use position-specific scores for amino acids or nucleotides. The algorithm continues to incorporate information from newly identified sequences to increase its motif analysis capabilities.

#### IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; and Ausubel, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar.

The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of Northern analysis are reported as a list of libraries in which the transcript encoding PRAEP occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

## V. Extension of PRAEP Encoding Polynucleotides

The nucleic acid sequence of Incyte Clone 1813005 was used to design oligonucleotide primers for extending a partial nucleotide sequence to full length. One primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 primer analysis software (National Biosciences, Plymouth, MN), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR

kit (Applied Biosystems) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the Peltier thermal cycler (PTC200; M.J. Research, Watertown, MA), beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

|    |         |   |
|----|---------|---|
| 5  | Step 1  | 94° C for 1 min (initial denaturation)                |
|    | Step 2  | 65° C for 1 min                                       |
|    | Step 3  | 68° C for 6 min                                       |
|    | Step 4  | 94° C for 15 sec                                      |
|    | Step 5  | 65° C for 1 min                                       |
| 10 | Step 6  | 68° C for 7 min                                       |
|    | Step 7  | Repeat steps 4 through 6 for an additional 15 cycles  |
|    | Step 8  | 94° C for 15 sec                                      |
|    | Step 9  | 65° C for 1 min                                       |
|    | Step 10 | 68° C for 7:15 min                                    |
| 15 | Step 11 | Repeat steps 8 through 10 for an additional 12 cycles |
|    | Step 12 | 72° C for 8 min                                       |
|    | Step 13 | 4° C (and holding)                                    |

A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQUICK gel purification kit (QIAGEN Inc.), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13  $\mu$ l of ligation buffer, 1  $\mu$ l T4-DNA ligase (15 units) and 1  $\mu$ l T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent *E. coli* cells (in 40  $\mu$ l of appropriate media) were transformed with 3  $\mu$ l of ligation mixture and cultured in 80  $\mu$ l of SOC medium. (See, e.g., Sambrook, *supra*, Appendix A, p. 2.) After incubation for one hour at 37° C, the *E. coli* mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, *supra*, Appendix A, p. 1) containing carbenicillin (2x carb). The following day, several colonies were randomly picked from each plate and cultured in 150  $\mu$ l of liquid LB/2x Carb medium placed in an individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5  $\mu$ l of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5  $\mu$ l from each sample was transferred into a PCR array.

For PCR amplification, 18  $\mu$ l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

|    |        |  |
|----|--------|--|
| 5  | Step 1 | 94° C for 60 sec                                     |
|    | Step 2 | 94° C for 20 sec                                     |
|    | Step 3 | 55° C for 30 sec                                     |
|    | Step 4 | 72° C for 90 sec                                     |
|    | Step 5 | Repeat steps 2 through 4 for an additional 29 cycles |
| 10 | Step 6 | 72° C for 180 sec                                    |
|    | Step 7 | 4° C (and holding)                                   |

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequence of SEQ ID NO:2 is used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

## VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:2 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 primer analysis software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham, Chicago, IL), and T4 polynucleotide kinase (DuPont NEN, Boston, MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine resin column (Pharmacia). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, XbaI, or Pvu II (DuPont NEN, Boston, MA).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (NYTRAN PLUS, Schleicher & Schuell, Durham, NH).

Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR film (Kodak, Rochester, NY) is exposed to the blots, hybridization patterns are compared visually.

5

## VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE. Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; and Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

## VIII. Complementary Polynucleotides

Sequences complementary to the PRAEP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PRAEP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the

same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 primer analysis software and the coding sequence of PRAEP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding  
5 sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PRAEP-encoding transcript.

## IX. Expression of PRAEP

Expression and purification of PRAEP is achieved using bacterial or virus-based  
10 expression systems. For expression of PRAEP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed  
15 into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PRAEP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PRAEP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA  
20 encoding PRAEP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus.  
25 (See Engelhard, E. K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PRAEP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell  
30 lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification

of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Pharmacia). Following purification, the GST moiety can be proteolytically cleaved from PRAEP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak, Rochester, NY). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN Inc, Chatsworth, CA). Methods for protein expression and purification are discussed in Ausubel, F. M. et al. (1995 and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch 10, 16. Purified PRAEP obtained by these methods can be used directly in the following activity assay.

#### **X. Demonstration of PRAEP Activity**

An assay for PRAEP activity measures the extent of transformation of NIH3T3 mouse fibroblast cells in which PRAEP is overexpressed. cDNA encoding PRAEP is subcloned into an appropriate mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. This construct is transfected into NIH3T3 cells using methods known in the art. Transfected cells are assessed for the following quantifiable properties characteristic of oncogenically transformed cells: growth in culture to high density associated with loss of contact inhibition, growth in suspension or in soft agar, lowered serum requirements, and ability to induce tumors when injected into immunodeficient (nu/nu) mice. The activity of PRAEP is proportional to the extent of transformation of NIH3T3 cells.

#### **XI. Functional Assays**

PRAEP function is assessed by expressing the sequences encoding PRAEP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV.SPORT plasmid (Life Technologies, Gaithersburg, MD) and pCR 3.1 plasmid (Invitrogen, Carlsbad, CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome

formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g.,

5 Green Fluorescent Protein (GFP) (Clontech, Palo Alto, CA), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate properties such as their apoptotic state. FCM detects and quantifies the uptake of fluorescent molecules that

10 diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the

15 binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York, NY.

The influence of PRAEP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PRAEP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to

20 conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success, NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PRAEP and other genes of interest can be analyzed by Northern analysis or microarray

25 techniques.

## **XII. Production of PRAEP Specific Antibodies**

PRAEP substantially purified using PAGE electrophoresis (see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

30 Alternatively, the PRAEP amino acid sequence is analyzed using LASERGENE



software (DNASTAR Inc.) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel supra, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an Applied Biosystems peptide synthesizer Model 431A using fmoc-chemistry and coupled to KLH (Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

### **XIII. Purification of Naturally Occurring PRAEP Using Specific Antibodies**

Naturally occurring or recombinant PRAEP is substantially purified by immunoaffinity chromatography using antibodies specific for PRAEP. An immunoaffinity column is constructed by covalently coupling anti-PRAEP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Pharmacia). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PRAEP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PRAEP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PRAEP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PRAEP is collected.

### **XIV. Identification of Molecules Which Interact with PRAEP**

PRAEP, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PRAEP,

washed, and any wells with labeled PRAEP complex are assayed. Data obtained using different concentrations of PRAEP are used to calculate values for the number, affinity, and association of PRAEP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

0986356-052401

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Bandman, Olga  
Corley, Neil C.  
Guegler, Karl J.  
Lal, Preeti
- (ii) TITLE OF THE INVENTION: PROSTATE-ASSOCIATED ETS PROTEIN
- (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
  - (B) STREET: 3174 Porter Dr.
  - (C) CITY: Palo Alto
  - (D) STATE: CA
  - (E) COUNTRY: USA
  - (F) ZIP: 94304
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Diskette
  - (B) COMPUTER: IBM Compatible
  - (C) OPERATING SYSTEM: DOS
  - (D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: To Be Assigned
  - (B) FILING DATE: Filed Herewith
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Billings, Lucy J.
  - (B) REGISTRATION NUMBER: 36,749
  - (C) REFERENCE/DOCKET NUMBER: PF-0501 US
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 650-855-0555
  - (B) TELEFAX: 650-845-4166
  - (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 335 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
  - (A) LIBRARY: PROSTUT12
  - (B) CLONE: 1813005

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Gly Ser Ala Ser Pro Gly Leu Ser Ser Val Ser Pro Ser His Leu  
1 5 10 15

Leu Leu Pro Pro Asp Thr Val Ser Arg Thr Gly Leu Glu Lys Ala Ala  
 20 25 30  
 Ala Gly Ala Val Gly Leu Glu Arg Asp Trp Ser Pro Ser Pro Pro  
 35 40 45  
 Ala Thr Pro Glu Gln Gly Leu Ser Ala Phe Tyr Leu Ser Tyr Phe Asp  
 50 55 60  
 Met Leu Tyr Pro Glu Asp Ser Ser Trp Ala Ala Lys Ala Pro Gly Ala  
 65 70 75 80  
 Ser Ser Arg Glu Glu Pro Pro Glu Glu Pro Glu Gln Cys Pro Val Ile  
 85 90 95  
 Asp Ser Gln Ala Pro Ala Gly Ser Leu Asp Leu Val Pro Gly Gly Leu  
 100 105 110  
 Thr Leu Glu Glu His Ser Leu Glu Gln Val Gln Ser Met Val Val Gly  
 115 120 125  
 Glu Val Leu Lys Asp Ile Glu Thr Ala Cys Lys Leu Leu Asn Ile Thr  
 130 135 140  
 Ala Asp Pro Met Asp Trp Ser Pro Ser Asn Val Gln Lys Trp Leu Leu  
 145 150 155 160  
 Trp Thr Glu His Gln Tyr Arg Leu Pro Pro Met Gly Lys Ala Phe Gln  
 165 170 175  
 Glu Leu Ala Gly Lys Glu Leu Cys Ala Met Ser Glu Glu Gln Phe Arg  
 180 185 190  
 Gln Arg Ser Pro Leu Gly Gly Asp Val Leu His Ala His Leu Asp Ile  
 195 200 205  
 Trp Lys Ser Ala Ala Trp Met Lys Glu Arg Thr Ser Pro Gly Ala Ile  
 210 215 220  
 His Tyr Cys Ala Ser Thr Ser Glu Glu Ser Trp Thr Asp Ser Glu Val  
 225 230 235 240  
 Asp Ser Ser Cys Ser Gly Gln Pro Ile His Leu Trp Gln Phe Leu Lys  
 245 250 255  
 Glu Leu Leu Leu Lys Pro His Ser Tyr Gly Arg Phe Ile Arg Trp Leu  
 260 265 270  
 Asn Lys Glu Lys Gly Ile Phe Lys Ile Glu Asp Ser Ala Gln Val Ala  
 275 280 285  
 Arg Leu Trp Gly Ile Arg Lys Asn Arg Pro Ala Met Asn Tyr Asp Lys  
 290 295 300  
 Leu Ser Arg Ser Ile Arg Gln Tyr Tyr Lys Lys Gly Ile Ile Arg Lys  
 305 310 315 320  
 Pro Asp Ile Ser Gln Arg Leu Val Tyr Gln Phe Val His Pro Ile  
 325 330 335

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1905 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PROSTUT12  
 (B) CLONE: 1813005

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| CTCGAGCCGC | GGCTGTCTGA | CTTCCTCCCA | GCACATTCCT | GCACTCTGCC | GTGTCCACAC | 60  |
| TGCCCCACAG | ACCCAGTCCT | CCAAGCCTGC | TGCCAGCTCC | CTGCAAGCCC | CTCAGGTTGG | 120 |
| GCCTTGCCAC | GGTGCCAGCA | GGCAGCCCTG | GGCTGGGGGT | AGGGGACTCC | CTACAGGCAC | 180 |
| GCAGCCCTGA | GACCTCAGAG | GGCCACCCCT | TGAGGGTGGC | CAGGCCCCCA | GTGGCCAACC | 240 |
| TGAGTGCTGC | CTCTGCCACC | AGCCCTGCTG | GCCCCTGGTT | CCGCTGGCCC | CCCAGATGCC | 300 |
| TGGCTGAGAC | ACGCCAGTGG | CCTCAGCTGC | CCACACCTCT | TCCCGGCCCC | TGAAGTTGGC | 360 |
| ACTGCAGCAG | ACAGCTCCCT | GGGCACCAGG | CAGCTAACAG | ACACAGCCGC | CAGCCCAAAC | 420 |
| AGCAGCGGCA | TGGGCAGCGC | CAGCCCGGGT | CTGAGCAGCG | TATCCCCCAG | CCACCTCCTG | 480 |
| CTGCCCCCCG | ACACGGTGTG | GCGGACAGGC | TTGGAGAAGG | CGGCAGCGGG | GGCAGTGGGT | 540 |
| CTCGAGAGAC | GGGACTGGAG | TCCAGTCCA  | CCCGCCACGC | CCGAGCAGGG | CCTGTCCGCC | 600 |
| TTCTACCTCT | CCTACTTTGA | CATGCTGTAC | CCTGAGGACA | GCAGCTGGGC | AGCCAAGGCC | 660 |

CCTGGGGCCA GCAGTCGGGA GGAGCCACCT GAGGAGCCTG AGCAGTGCCC GGTCATTGAC 720  
 AGCCAAGCCC CAGCGGGCAG CCTGGACTTG GTGCCC GGCTGACCTT GGAGGAGCAC 780  
 TCGCTGGAGC AGGTGCAGTC CATGGTGGTG GCGAAGTGC TCAAGGACAT CGAGACGGCC 840  
 TGCAAGCTGC TCAACATCAC CGCAGATCCC ATGGACTGGA GCCCCAGCAA TGTGCAGAAG 900  
 TGGCTCCTGT GGACAGAGCA CCAATACCGG CTGCCCCCA TGGGCAAGGC CTTCCAGGAG 960  
 CTGGCGGGCA AGGAGCTGTG CGCCATGTCTG GAGGAGCAGT TCCGCCAGCG CTCGCCCCTG 1020  
 GGTGGGGATG TGCTGCACGC CCACCTGGAC ATCTGGAAGT CAGCGGCCTG GATGAAAGAG 1080  
 CGGACTTCAC CTGGGGCGAT TCACTACTGT GCCTCGACCA GTGAGGAGAG CTGGACCGAC 1140  
 AGCGAGGTGG ACTCATCATG CTCCGGGCAG CCCATCCACC TGTGGCAGTT CCTCAAGGAG 1200  
 TTGCTACTCA AGCCCCACAG CTATGGCCGC TTCATTAGGT GGCTCAACAA GGAGAAGGGC 1260  
 ATCTTCAAAA TTGAGGACTC AGCCCAGGTG GCCCCGCTGT GGGGCATCCG CAAGAACCGT 1320  
 CCCGCCATGA ACTACGACAA GCTGAGCCGC TCCATCCGCC AGTATTACAA GAAGGGCATC 1380  
 ATCCGGAAGC CAGACATCTC CCAGCGCCTC GTCTACCAGT TCGTGCACCC CATCTGAGTG 1440  
 CCTGGCCCAG GGCCTGAAAC CCGCCCTCAG GGGCCTCTCT CCTGCCTGCC CTGCCTCAGC 1500  
 CAGGCCCTGA GATGGGGGAA AACGGGCAGT CTGCTCTGCT GCTCTGACCT TCCAGAGCCC 1560  
 AAGGTCAGGG AGGGGCAACC AACTGCCCCA GGGGGATATG GGTCTCTGCT GGCCTTCGGG 1620  
 ACCCTGGGGC AGGGGTGCTT CCTCCTCAGG CCCAGCTGCT CCCCTGGAGG ACAGAGGGAG 1680  
 ACAGGGCTGC TCCCCAACAC CTGCCTCTGA CCCCAGCATT TCCAGAGCAG AGCCTACAGA 1740  
 AGGGCAGTGA CTCGACAAAG GCCACAGGCA GTCCAGGCCT CTCTCTGCTC CATCCCCCTG 1800  
 CCTCCCATTG TGCACCACAC CTGGCATGGT GCAGGGAGAC ATCTGCACCC CTGAGTTGGG 1860  
 CAGCCAGGAG TGCCCCCGGG AATGGATAAT AAAGATACTA GGCGC 1905

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 114 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 157196

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Thr | Asn | Ala | Ser | Asn | Gly | Gly | Thr | Ala | Thr | Val | Lys | Arg | Pro | Asn | Gly |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Gly | Arg | Thr | Gly | Gly | Gly | Gly | Ser | His | Ile | His | Leu | Trp | Gln | Phe | Leu |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Lys | Glu | Leu | Leu | Ala | Ser | Pro | Gln | Val | Asn | Gly | Thr | Ala | Ile | Arg | Trp |
|     |     | 35  |     |     |     | 40  |     |     |     |     |     | 45  |     |     |     |
| Ile | Asp | Arg | Ser | Lys | Gly | Ile | Phe | Lys | Ile | Glu | Asp | Ser | Val | Arg | Val |
|     | 50  |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |     |
| Ala | Lys | Leu | Trp | Gly | Arg | Arg | Lys | Asn | Arg | Pro | Ala | Met | Asn | Tyr | Asp |
| 65  |     |     |     |     | 70  |     |     |     | 75  |     |     |     |     | 80  |     |
| Lys | Leu | Ser | Arg | Ser | Ile | Arg | Gln | Tyr | Tyr | Lys | Lys | Gly | Ile | Met | Lys |
|     |     |     | 85  |     |     |     |     | 90  |     |     |     |     |     | 95  |     |
| Lys | Thr | Glu | Arg | Ser | Gln | Arg | Leu | Val | Tyr | Gln | Phe | Cys | His | Pro | Tyr |
|     |     |     | 100 |     |     |     | 105 |     |     |     |     |     | 110 |     |     |
| Ser | Gln |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 452 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 511283

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ser | Glu | Thr | Pro | Ala | Gln | Cys | Ser | Ile | Lys | Gln | Glu | Arg | Ile | Ser | 1   | 5   | 10  | 15  |
| Tyr | Thr | Pro | Pro | Glu | Ser | Pro | Val | Pro | Ser | Tyr | Ala | Ser | Ser | Thr | Pro | 20  | 25  | 30  |     |
| Leu | His | Val | Pro | Val | Pro | Arg | Ala | Leu | Arg | Met | Glu | Glu | Asp | Ser | Ile | 35  | 40  | 45  |     |
| Arg | Leu | Pro | Ala | His | Leu | Arg | Leu | Gln | Pro | Ile | Tyr | Trp | Ser | Arg | Asp | 50  | 55  | 60  |     |
| Asp | Val | Ala | Gln | Trp | Leu | Lys | Trp | Ala | Glu | Asn | Glu | Phe | Ser | Leu | Arg | 65  | 70  | 75  | 80  |
| Pro | Ile | Asp | Ser | Asn | Thr | Phe | Glu | Met | Asn | Gly | Lys | Ala | Leu | Leu | Leu | 85  | 90  | 95  |     |
| Leu | Thr | Lys | Glu | Asp | Phe | Arg | Tyr | Arg | Ser | Pro | His | Ser | Gly | Asp | Val | 100 | 105 | 110 |     |
| Leu | Tyr | Glu | Leu | Leu | Gln | His | Ile | Leu | Lys | Gln | Arg | Lys | Pro | Arg | Ile | 115 | 120 | 125 |     |
| Leu | Phe | Ser | Pro | Phe | Phe | His | Pro | Gly | Asn | Ser | Ile | His | Thr | Gln | Pro | 130 | 135 | 140 |     |
| Glu | Val | Ile | Leu | His | Gln | Asn | His | Glu | Glu | Asp | Asn | Cys | Val | Gln | Arg | 145 | 150 | 155 | 160 |
| Thr | Pro | Arg | Pro | Ser | Val | Asp | Asn | Val | His | His | Asn | Pro | Pro | Thr | Ile | 165 | 170 | 175 |     |
| Glu | Leu | Leu | His | Arg | Ser | Arg | Ser | Pro | Ile | Thr | Thr | Asn | His | Arg | Pro | 180 | 185 | 190 |     |
| Ser | Pro | Asp | Pro | Glu | Gln | Arg | Pro | Leu | Arg | Ser | Pro | Leu | Asp | Asn | Met | 195 | 200 | 205 |     |
| Ile | Arg | Arg | Leu | Ser | Pro | Ala | Glu | Arg | Ala | Gln | Gly | Pro | Arg | Pro | His | 210 | 215 | 220 |     |
| Gln | Glu | Asn | Asn | His | Gln | Glu | Ser | Tyr | Pro | Leu | Ser | Val | Ser | Pro | Met | 225 | 230 | 235 | 240 |
| Glu | Asn | Asn | His | Cys | Pro | Ala | Ser | Ser | Glu | Ser | His | Pro | Lys | Pro | Ser | 245 | 250 | 255 |     |
| Ser | Pro | Arg | Gln | Glu | Ser | Thr | Arg | Val | Ile | Gln | Leu | Met | Pro | Ser | Pro | 260 | 265 | 270 |     |
| Ile | Met | His | Pro | Leu | Ile | Leu | Asn | Pro | Arg | His | Ser | Val | Asp | Phe | Lys | 275 | 280 | 285 |     |
| Gln | Ser | Arg | Leu | Ser | Glu | Asp | Gly | Leu | His | Arg | Glu | Gly | Lys | Pro | Ile | 290 | 295 | 300 |     |
| Asn | Leu | Ser | His | Arg | Glu | Asp | Leu | Ala | Tyr | Met | Asn | His | Ile | Met | Val | 305 | 310 | 315 | 320 |
| Ser | Val | Ser | Pro | Pro | Glu | Glu | His | Ala | Met | Pro | Ile | Gly | Arg | Ile | Ala | 325 | 330 | 335 |     |
| Asp | Cys | Arg | Leu | Leu | Trp | Asp | Tyr | Val | Tyr | Gln | Leu | Leu | Ser | Asp | Ser | 340 | 345 | 350 |     |
| Arg | Tyr | Glu | Asn | Phe | Ile | Arg | Trp | Glu | Asp | Lys | Glu | Ser | Lys | Ile | Phe | 355 | 360 | 365 |     |
| Arg | Ile | Val | Asp | Pro | Asn | Gly | Leu | Ala | Arg | Leu | Trp | Gly | Asn | His | Lys | 370 | 375 | 380 |     |
| Asn | Arg | Thr | Asn | Met | Thr | Tyr | Glu | Lys | Met | Ser | Arg | Ala | Leu | Arg | His | 385 | 390 | 395 | 400 |
| Tyr | Tyr | Lys | Leu | Asn | Ile | Ile | Arg | Lys | Glu | Pro | Gly | Gln | Arg | Leu | Leu | 405 | 410 | 415 |     |
| Phe | Arg | Phe | Met | Lys | Thr | Pro | Asp | Glu | Ile | Met | Ser | Gly | Arg | Thr | Asp | 420 | 425 | 430 |     |
| Arg | Leu | Glu | His | Leu | Glu | Ser | Gln | Glu | Leu | Asp | Glu | Gln | Ile | Tyr | Gln | 435 | 440 | 445 |     |
| Glu | Asp | Glu | Cys |     |     |     |     |     |     |     |     |     |     |     |     | 450 |     |     |     |